УДК 534.4

ИССЛЕДОВАНИЕ СРЕДСТВ И ТЕХНОЛОГИЙ СИНТЕЗА РЕЧИ

Руденко А.И., Никитюк К.В.
к.т.н., доц., каф. ИС Николенко А.А.
Одесский Национальный Политехнический Университет, УКРАИНА

АННОТАЦИЯ. В данном исследовании рассмотрены проблемы, связанные с синтезом устной речи. В частности рассматривается одна из доступных моделей называемая «unit selection», её особенности, преимущества и недостатки. Рассмотрены подходы и способы улучшения данной модели, а так же направления для дальнейших исследований.

Введение. Синтезом речи на сегодняшний день называют технологию, способную преобразовывать текстовую информацию в устную речь. С развитием компьютеров эта технология становится всё более актуальной, и каждый день совершенствуется. Данное исследование является актуальным потому, что в настоящее время использование подобных систем является очень востребованным. Задачей данного исследования является изучение способов улучшения представленной модели синтеза речи. В данном исследовании не будут рассматриваться задачи фонематического анализа текста.

Основная часть. Есть несколько принципиально разных технологий синтеза речи, и в большинстве современных систем используется конкатенативный синтез методом «unit selection». Заранее записанный образец голоса режется на определенные составные элементы (например, контекстно-зависимые фонемы), из которых составляется речевая база. Затем любые нужные слова собираются из отдельных юнитов. Получается правдоподобная имитация человеческого голоса, но воспринимать его тяжело — скачет тембр, возникают неестественные интонации и резкие переходы на стыках отдельных юнитов. Особенно это заметно при озвучивании длинного связного текста. Качество такой системы можно повышать, увеличивая объём речевой базы, но это долгий и кропотливый труд, требующий привлечения профессионального и очень терпеливого диктора. И полнота базы всегда остаётся узким местом системы. Следует также отметить возможность использования и более крупных речевых элементов т. к. слоги и слова.

В данном исследовании для заполнения речевой базы был использован материал полученный не от живого диктора, а синтезированный с помощью Yandex SpeechKit. В качестве юнитов были выбраны отдельные звуки букв от «а» до «я», за исключением беззвучных «ъ» и «ь», звуки возможных буквенных комбинаций от «аа» до «яя», в данном случае исключений было намного больше, список которых был позаимствован из исследования [1]. Таким образом из $33+33^2=1122$ юнитов остались только $31+868=899$ юнитов, что позволило немного уменьшить размер речевой базы. Кроме того было решено добавить в примерно 300-500 часто употребляемых слов встречающихся в текстах рассматриваемой тематики, в данном случае это сказки для детей 3-6 лет. Для получения списка этих слов были проанализированы более 50 книг полученных на веб-ресурсе [2].

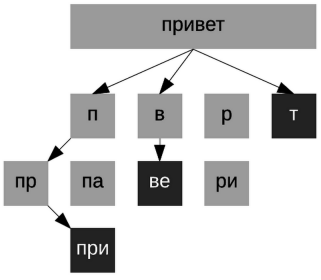

Рис. *1* - Дерево подбора необходимых юнитов

Для составления требуемого слова из базы данных берутся необходимые речевые части которые впоследствии объединяются во едино. В данный момент правила подбора необходимых юнитов достаточно просты, но в последствии могут быть улучшены.

В ходе проведения данного исследования удалось выделить следующие возможные способы улучшения качества работы данной модели:

— увеличение количества юнитов и их типов хранимых в речевой базе данных, чем больше будет разнообразие доступных элементов, тем более точно можно будет воспроизвести заданное слово;

— использование более продвинутых алгоритмов выбора элементов из речевой базы, ведь даже при наличии обширной базы плохой алгоритм выбора юнитов может значительно снизить общее качество системы за счет того, что он попросту будет выбирать плохие комбинации элементов;

— дополнительная обработка результирующего сигнала, т. к. на стыках элементов практически гарантированно будут слышны перепады тона и возможно некоторые шумы.

Преимущества данной модели заключаются в следующем:

— высокая четкость воспроизводимой речи, при этом следует отметить, что это не говорит о том, что речь звучит очень натурально, но при этом легко понять то, о чём говориться;

— при необходимости озвучивать достаточно ограниченный набор фраз, база данных может быть очень компактной;

Недостатки данной модели заключаются в следующем:

— для обеспечения высокой точности воспроизведения необходима очень большая база данных [3], создание которой является крайне трудоёмким процессом который в большинстве случаев выполняется людьми вручную [4];

— трудно управлять интонацией и темпом речи без значительных искажений натуральности [5][6];

В направления для дальнейших исследований следует выделить следующее:

— методы и технологии для автоматической нарезки дикторской речи на звуковые единицы, здесь могут пригодится технологии распознания речи и нейронные сети [6];

— алгоритмы и методы позволяющие сгладить неровности склейки элементов речи.

Выводы. В результате исследования можно сделать следующее заключение: данная модель хорошо подходит для решения поставленной задачи, а именно озвучивания детских сказок, т. к. предоставляет возможность получить более менее натуральный голос с понятной дикцией, к сожалению придётся пожертвовать звучностью синтезируемого голоса т. к. будут слышны искажения на стыках элементов. Кроме того очевидно что речевая база данных потребует большого количества усилий от разработчиков, т. к. качество системы будет напрямую зависеть от качества данных. Касательно алгоритмов следует отметить, что хоть они также являются важной частью системы но они не критичны, и по больше части могут быть достаточно простыми, чего нельзя сказать о пост-обработке сигнала после склейки элементов, что может вызнать трудности в реализации.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Невозможные буквосочетания в русском языке [Електронний ресурс]. — режим доступу: URL: http://vl2d.livejournal.com/21053.html

2. Сказки для 3-6 лет [Електронний ресурс]. — режим доступу: URL: http://skazochki.info/skazki-dlya/skazki-dlya-3-6-let

3. Синтез речи [Електронний ресурс]. — режим доступу: URL: http://voicenavigator.ru/technology-synthesis

4. Синтез речи [Електронний ресурс]. — режим доступу: URL: https://ru.wikipedia.org/wiki/Синтез_речи

5. Преобразование произвольной текстовой информации в речь [Електронний ресурс]. — режим доступу: URL: http://www.speetech.by/technologies/tts

6. Синтез и распознавание речи [Електронний ресурс]. — режим доступу: URL: http://www.frolov-lib.ru/books/hi/ch07.html